Assessing Systematic Impacts of PBL Schemes on Storm Evolution in the NOAA Warn-on-Forecast System

COREY K. POTVIN,[a,b] PATRICK S. SKINNER,[c,a,b] KIMBERLY A. HOOGEWIND,[c,a] MICHAEL C. CONIGLIO,[a,b] JEREMY A. GIBBS,[c,a] ADAM J. CLARK,[a,b] MONTGOMERY L. FLORA,[b,c,a] ANTHONY E. REINHART,[c,a] JACOB R. CARLEY,[d] AND ELIZABETH N. SMITH[a]

[a] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[b] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[c] *Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*
[d] *NOAA/NWS/NCEP, Environmental Modeling Center, College Park, Maryland*

(Manuscript received 27 November 2019, in final form 23 March 2020)

ABSTRACT

The NOAA Warn-on-Forecast System (WoFS) is an experimental rapidly updating convection-allowing ensemble designed to provide probabilistic operational guidance on high-impact thunderstorm hazards. The current WoFS uses physics diversity to help maintain ensemble spread. We assess the systematic impacts of the three WoFS PBL schemes—YSU, MYJ, and MYNN—using novel, object-based methods tailored to thunderstorms. Very short forecast lead times of 0–3 h are examined, which limits phase errors and thereby facilitates comparisons of observed and model storms that occurred in the same area at the same time. This evaluation framework facilitates assessment of systematic PBL scheme impacts on storms and storm environments. Forecasts using all three PBL schemes exhibit overly narrow ranges of surface temperature, dewpoint, and wind speed. The surface biases do not generally decrease at later forecast initialization times, indicating that systematic PBL scheme errors are not well mitigated by data assimilation. The YSU scheme exhibits the least bias of the three in surface temperature and moisture and in many sounding-derived convective variables. Interscheme environmental differences are similar both near and far from storms and qualitatively resemble the differences analyzed in previous studies. The YSU environments exhibit stronger mixing, as expected of nonlocal PBL schemes; are slightly less favorable for storm intensification; and produce correspondingly weaker storms than the MYJ and MYNN environments. On the other hand, systematic interscheme differences in storm morphology and storm location forecast skill are negligible. Overall, the results suggest that calibrating forecasts to correct for systematic differences between PBL schemes may modestly improve WoFS and other convection-allowing ensemble guidance at short lead times.

1. Introduction

One challenge of objective model evaluation is to focus upon atmospheric features of greatest interest rather than treating all fields and grid points equally. Object-based methods address this challenge by extracting features from the model state and diagnosing operationally and/or scientifically important attributes (e.g., Wolff et al. 2014). Object-based methods are particularly well suited to verifying convection-allowing model (CAM) analyses and forecasts since discrete features, such as thunderstorms, and their attributes are usually of primary interest (e.g., Johnson et al. 2013; Stratman and Brewster 2017; Jones et al. 2018; Skinner et al. 2018; Potvin et al. 2019; Adams-Selin et al. 2019; Flora et al. 2019; Lawson et al. 2020, manuscript submitted to *Mon. Wea Rev.*). A second challenge for model evaluation is distilling biases and systematic differences (e.g., between forecasts generated using different model configurations) from many diverse cases. Composite analysis techniques provide an objective way to illuminate such effects and then communicate them to forecast users and model developers.

Potvin et al. (2019, hereafter P19) presented object-based, composite analysis techniques for evaluating and comparing CAM next-day forecasts. That work utilized the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018), a major feature of the 2016–19 NOAA Hazardous Weather Testbed Spring Forecasting Experiments (SFEs; Kain et al. 2003;

Corresponding author: Dr. Corey K. Potvin, corey.potvin@noaa.gov

DOI: 10.1175/MWR-D-19-0389.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

Clark et al. 2012; Gallo et al. 2017). The present study extends the framework of P19 to evaluating systematic forecast impacts of the three planetary boundary layer (PBL) schemes used within the NOAA Warn-on-Forecast (WoF; Stensrud et al. 2009, 2013) System (WoFS), an experimental CAM ensemble that is slated to be operationalized as part of the Unified Forecast System to provide short-range probabilistic guidance for thunderstorm hazards. The much shorter forecast lead times examined in the present study (0–3 h) than in P19 (18–26 h) greatly increases the number of cases where all the considered models—in this study, WoFS members using each of the three PBL schemes—produce a storm in proximity to the same observed storm. This allows us to directly compare simulated and observed storms that occurred in approximately the same location, time, and environment, and thereby isolate PBL scheme impacts on modeled storms and near-storm environments.

There are at least three motivations for assessing systematic impacts of parameterization schemes in multiphysics ensembles. First, knowledge of systematic physics scheme errors can inform the interpretation and postprocessing of ensemble forecasts (e.g., Johnson et al. 2011). For example, in generating probabilistic forecasts of storm intensity, ensemble members that use physics schemes that produce large storm intensity errors could be weighted less than other ensemble members. Second, knowledge of the deficiencies in current physics schemes can inform the development of new and improved schemes. Third, identifying and replacing underperforming physics schemes in an ensemble system may improve the accuracy of the ensemble. While considerable attention has been given to the behavior of different PBL schemes at 6–36-h lead times (e.g., Hu et al. 2010; Coniglio et al. 2013; Cohen et al. 2015; Burlingame et al. 2017), much less is known about PBL scheme impacts, particularly on forecasts of storms, at the $O(1)$ h lead times in the purview of WoF. To the authors' knowledge, Kerr et al. (2017) is the only study to have addressed this question.

The WoFS and the observational datasets used to verify the analyses and forecasts in this study are described in section 2. Procedures for extracting storm objects and model fields within the near-storm environment are detailed in section 3. In section 4, novel verification techniques are used to evaluate and compare forecast biases associated with the different WoFS PBL schemes. Additional systematic physics impacts on storm attributes and near-storm fields are examined in section 5. Finally, section 6 summarizes the major conclusions of this study and recommendations for future work.

2. WoFS and observational datasets

a. *WoFS configuration*

The WoFS is a rapidly updating ensemble data assimilation and prediction system designed to provide probabilistic forecast guidance for thunderstorm hazards including tornadoes, damaging winds, hail, heavy rainfall, and lightning. The WoFS comprises 36 ensemble members that use the Advanced Research version 3.8.1 of the Weather Research and Forecasting (WRF) Model dynamical core (ARW; Skamarock et al. 2008; Powers et al. 2017) with 3-km horizontal grid spacing and 51 vertical levels. Radar and satellite observations are assimilated every 15 min and conventional observations every hour using an ensemble Kalman filter. During the 2017 (2018) NOAA Hazardous Weather Testbed SFEs, the WoFS was run over a 750 km × 750 km (900 km × 900 km) domain whose daily location was determined in collaboration with the NOAA Storm Prediction Center. Initial and boundary conditions for each WoFS ensemble member are provided by the corresponding member of the 36-member High-Resolution Rapid Refresh (Benjamin et al. 2016) Ensemble (HRRRE; Dowell et al. 2016). In 2017–18, WoFS members were initialized from 1-h HRRRE forecasts valid at 1800 UTC. Each WoFS member uses one of three PBL parameterizations available in the WRF-ARW: the Yonsei University (YSU; Hong et al. 2006), Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 2002), or Mellor–Yamada–Nakanishi–Niino (MYNN; Nakanishi and Niino 2004, 2006) scheme; and one of two sets of radiation parameterizations: the Dudhia (1989) shortwave and Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave schemes, or the Rapid Radiative Transfer Model for Global (RRTMG; Iacono et al. 2008) longwave and shortwave schemes. This yields a total of six unique physics combinations among the ensemble members. In earlier work, we examined systematic impacts of the two pairs of radiation schemes and of the six PBL–radiation scheme combinations. Those preliminary analyses revealed the choice of PBL scheme has a much greater impact on forecasts than the choice of radiation schemes, consistent with Kerr et al. (2017); thus, for brevity, we will not discuss those experiments in this paper. All members use the RUC land surface model (Smirnova et al. 2016) and the NSSL two-moment microphysics scheme (Mansell et al. 2010). Only members 1–18 are used to generate free forecasts; these members use the same physics as members 19–36, respectively. Herein we analyze 0–3-h WoFS forecasts initialized hourly at 1900–0200 UTC over 40 days during the 2017 and 2018 SFEs.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

b. *MRMS products, ASOS observations, and soundings*

The NSSL Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) system is used to verify WoFS forecasts of storm location and morphological attributes. Model composite reflectivity (REFLCOMP) is verified with the MRMS REFLCOMP product that is computed by calculating the exponential inverse-distance-weighted average of the reflectivity from each contributing radar and then taking the column-maximum (Smith et al. 2016). The MRMS REFLCOMP is interpolated to the WoFS grid using the Cressman scheme with a 3-km radius of influence. Automated Surface Observing System (ASOS) observations of 2-m temperature (T_2M), 2-m dewpoint (TD_2M), and 10-m wind components u and v (U_10M, V_10M) are used to verify forecasts of the same variables and of 10-m wind speed (WSPD_10M) in the vicinity of storms. National Weather Service rawinsonde observations valid within the WoFS domains analyzed in this study are used to verify model profiles of temperature, dewpoint, mixing ratio, u, and v, along with several sounding-derived parameters commonly used in severe thunderstorm forecasting. The rawinsonde observations were obtained from the University of Wyoming sounding database (http://weather.uwyo.edu/upperair/sounding.html) via the Unidata Siphon Python package version 0.8 (May et al. 2017). Both the model and observed sounding-derived parameters are calculated using the Sounding and Hodograph Analysis and Research Program in Python (SHARPpy) package (Blumberg et al. 2017). Following Coniglio et al. (2013), we compare the observed soundings to model soundings valid one hour prior, since 1) radiosondes are typically launched nearly one hour before their nominally valid times, and 2) we are primarily interested in the lowest kilometer of the vertical error profiles.

Definitions of all variables examined in this study are provided in Table 1.

TABLE 1. Descriptions of the primary variables analyzed in this study. The lowest 100-mb layer was used to calculate mixed-layer (ML) variables. STP_{fixed} is computed as in Thompson et al. (2012) except using $SRH_{0–1}$ instead of SRH valid over the effective inflow layer.

Variable	Description
REFLCOMP	Composite reflectivity (dBZ)
$UH_{2–5}$, $UH_{0–2}$	Hourly maximum 2–5-/0–2-km updraft helicity ($m^2 s^{-2}$)
RAIN_1H	Hourly accumulated rainfall (in.)
WMAX	Column-maximum vertical velocity ($m s^{-1}$)
SBCAPE, MLCAPE	Surface-based/mixed-layer convective available potential energy ($J kg^{-1}$)
$SRH_{0–3}$, $SRH_{0–1}$	0–3-/0–1-km storm relative helicity ($m^2 s^{-2}$)
T_2M	2-m AGL temperature (°C)
TD_2M	2-m AGL dewpoint (°C)
SBCIN, MLCIN	Surface-based/mixed-layer convective inhibition ($J kg^{-1}$)
MLSTP	Mixed-layer significant tornado parameter
STP_{fixed}	Fixed-layer significant tornado parameter
SCP	Supercell composite parameter
VORTMAX	Column-maximum vertical vorticity below 2 km AGL (s^{-1})
MLLCL	Mixed-layer lifted condensation level (m)
U_10M, V_10M	10-m AGL zonal, meridional wind components ($m s^{-1}$)
WSPD_10M	10-m AGL wind speed ($m s^{-1}$)
MAXHAIL	Maximum hail diameter (in.) at surface from WRF-HAILCAST (Adams-Selin et al. 2019)
PBL_HGT	PBL top height (m AGL) computed as in Coniglio et al. (2013)

3. Storm object identification

Object-based methods focus evaluation upon discrete features (e.g., storms) of greatest interest to the user, and avoid traditional methods' unduly large penalty for phase errors. Such methods are particularly well suited to verifying and comparing WoF ensemble output, which is designed primarily to provide forecasters with guidance on near-term (0–3 h) evolution of potentially high-impact storms. As in P19, we extract both storm objects and, in the case of model storms, their near-storm environments (NSEs). Each NSE is a prescribed set of model fields within a 120 km × 120 km domain centered on the storm object centroid. The NSE domain is sized to encompass storm–environment interactions that can modulate storm intensity, not to represent the preconvective environment, and is consistent with the recommendations of Potvin et al. (2010). Most of our analysis of environmental variables is conducted within these NSEs since we are primarily interested in PBL scheme impacts within and near simulated storms.

The first step of the storm object[1] extraction is to identify regions of the MRMS and WoFS REFLCOMP fields exceeding prescribed thresholds. The MRMS REFLCOMP threshold is set to the 99.9th percentile of the set of MRMS REFLCOMP values compiled over all the forecasts used in this study. The WoFS REFLCOMP threshold for each PBL scheme is computed similarly but across all ensemble members using that scheme. The resulting MRMS REFLCOMP threshold is 40.5 dBZ

[1] Object identification, extraction, and characterization in this study were performed using the Python Scikit-image library (Van der Walt et al. 2014).

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

and the WoFS REFLCOMP thresholds range over 44.0–44.3 dBZ. Using percentile thresholds accounts for the systematic difference between the MRMS and WoFS REFLCOMP.

Preliminary storm objects identified using the REFLCOMP thresholding procedure are then merged into a single object if their boundaries lie within 10 km of each other. This step prevents mesoscale convective systems (MCSs) with localized weaknesses in their intense convection from being misidentified as multiple discrete storms. Next, objects with area <12 grid cells (108 km^2) are discarded since very small objects are less likely to correspond to the intense, organized storms which are the focus of this study. Finally, to exclude MCSs, storm objects with length >75 km or area >2500 km^2 are discarded.[2] Restricting the analysis to discrete storms avoids the difficulty of tailoring analysis methods to very different storm modes and facilitates interpretation of the results. It would be valuable, however, to extend our methodology to MCSs in future work to determine whether systematic PBL scheme impacts vary between MCSs and more discrete storms.

In assessing systematic PBL scheme impacts, it would not make sense to compare two ensemble member forecasts within a region where only one member contains a storm, since the differences between the two forecasts could arise largely from the presence of a storm (and attendant storm–environment interactions) in one member and the absence of that storm in the other, and not necessarily from the use of different PBL schemes. This consideration together with our objective of evaluating PBL scheme impacts on model storms and NSEs motivates the development of a framework in which inter-PBL-scheme comparisons are restricted to storm-object-containing regions. Owing to the short (0–3-h) forecast lead times in the present study, there are numerous instances where model storms simulated with each of the three WoFS PBL schemes occur in proximity to an observed (MRMS) storm. To exploit this property of the WoFS forecasts, we cycle through each MRMS storm (object) and search for model storms whose centroid lies within 40 km of the MRMS storm centroid at the same valid time. The first such storm identified for each PBL scheme is provisionally retained. If such a storm is not identified for all three PBL schemes, then the MRMS storm and provisionally retained WoFS storms are discarded. Otherwise, the four (MRMS, YSU, MYNN, and MYJ) storm objects, referred to hereafter as a "storm tetrad" (Fig. 1), are retained and their correspondence to one another is utilized in subsequent analysis.[3] This selection procedure yields 4398 storm tetrads that are then used for all analysis in this study except for the storm object matching verification (section 4a), which uses the full set of WoFS and MRMS storm objects (6311 MRMS storms and approximately 10 000 storms for each WoFS member).

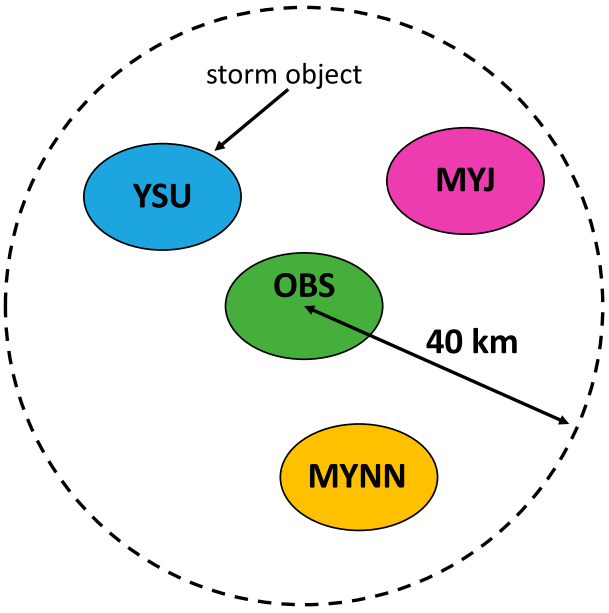

FIG. 1. Schematic of a storm tetrad. Each model storm object centroid lies within 40 km of the observed storm centroid at the same forecast valid time.

Restricting the analysis to these storm tetrads ensures that the analyzed mean interscheme differences arise primarily from systematic PBL scheme impacts, and are not substantially biased by sampling errors associated with uneven representation of convective scenarios among storms simulated with different PBL schemes. Analyzing differences between the three model storms within each tetrad rather than between model storms randomly drawn from each of the three unordered sets of storm objects also reduces the variance in the computed inter-PBL-scheme differences, and therefore the uncertainty in the mean interscheme differences. Finally, anchoring the storm tetrads on observed storms focuses the analysis on model storms that have been well

[2] Storm object length is computed by Scikit-image, and is the length of the major axis of the ellipse having the same normalized second central moments as the storm object.

[3] Repeating our analysis using a 20-km proximity criterion for the storm tetrads produces similar mean inter-PBL-scheme differences as does the 40-km criterion. The uncertainty in the differences, however, is substantially increased since the number of storm tetrads is approximately halved with the stricter proximity criterion.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

constrained by the WoFS data assimilation and are therefore of greatest value to forecasters (as opposed to potentially spurious convection that should be given less credence). Further discussion of our storm tetrad methodology can be found in the appendix.

4. Physics impacts on forecast performance

a. Storm object matching

To assess PBL scheme impacts on the accuracy of analyzed and forecast storm locations, storm objects in each WoFS member and the MRMS storm objects are matched to one another at lead times of 0, 1, 2, and 3 h. The object matching is performed using the technique of Skinner et al. (2018), which was itself derived from the matching technique in the Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006a,b). The effective maximum allowable displacement between forecast and observed storm centroids and boundaries is 32 km and no allowance is made for timing errors. WoFS storm objects in MRMS REFLCOMP data voids are not included in the matching. Matched forecast objects are counted as hits, unmatched forecast objects as false alarms, and unmatched observed objects as misses. Probability of detection (POD), false alarm ratio (FAR), success ratio (SR; 1 − FAR), critical success index (CSI), and frequency bias are then computed from the totals of hits, misses, and false alarms for each ensemble member and plotted on a performance diagram (Roebber 2009; Fig. 2). None of the four verification statistics vary substantially between the PBL schemes (i.e., all three schemes produce similarly skillful 0–3-h forecasts of storm locations).[4]

b. Surface verification

All ASOS observations lying within extracted NSEs and collected within 2.5 min of the forecast valid time are compared to model values obtained by bilinear interpolation from the surrounding four grid points. Treating the observations as truth, forecast errors (interpolated-model-minus-observation differences) are computed for T_2M, TD_2M, U_10M, V_10M, and WSPD_10M and analyzed in several ways. First, to identify any systematic interscheme differences in the spatial configuration of the NSEs, the forecast bias (average error) for each PBL scheme is computed within 9 subdomains obtained by dividing the NSE domain into a 3 × 3 grid (Fig. 3). The confident lower bound for each bias is computed by bootstrapping (10 000 iterations) the forecast errors to produce a distribution of bias realizations.[5] Spatial gradients in bias were fairly similar across the three PBL schemes (Fig. 3). For example, TD_2M biases were lowest southwest of storm centroids and highest near and northeast of storm centroids in all three cases (Fig. 3b). It therefore appears that any systematic interscheme differences that may exist in the simulated storm–environment interactions are too small to qualitatively impact the spatial configuration of the NSE.

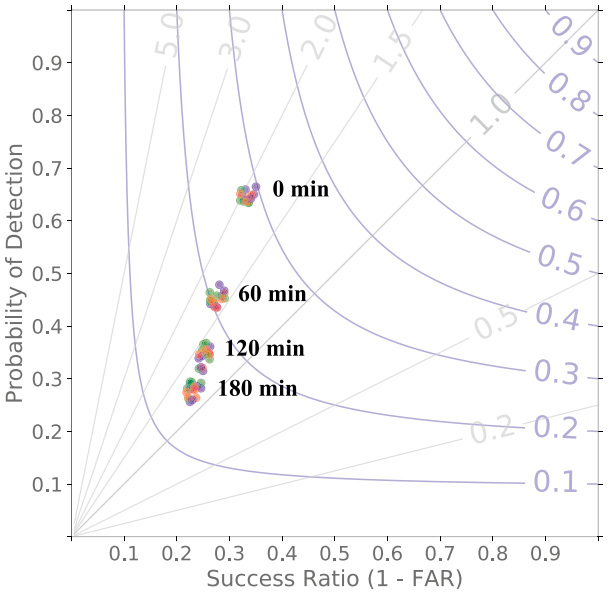

FIG. 2. Reflectivity object matching verification for each of the 18 WoFS forecast members, color-coded by PBL scheme (blue = YSU, red = MYJ, green = MYNN) and labeled by lead time. CSI and frequency bias are contoured in blue and gray, respectively.

Forecast error distributions for each PBL scheme are now examined and compared (Fig. 4). Using a similar bootstrapping procedure as for the previous analysis (Fig. 3), medians and confident lower bounds are computed for the forecast biases and the interscheme differences in bias. The YSU scheme consistently produced the warmest, driest NSEs, while the MYJ scheme produced the coolest, moistest NSEs (Figs. 4a,b). Our finding that the YSU scheme produces warmer, drier PBLs than the MYJ scheme in the 0–3-h forecasts is

[4] Additional experiments revealed that including MCSs in the verification and/or setting the maximum allowable temporal offset to 25 min substantially improved the POD, FAR, and CSI. In all three of those experiments, however, no single PBL scheme outperformed the others.

[5] If the 5th and 95th percentiles of this distribution have opposite signs (i.e., if the 90% confidence interval contains zero), then the bias is not statistically significantly different from zero, and the confident lower bound is therefore set to zero. If the 5th and 95th percentiles have the same sign, then the confident lower bound is set to the percentile value nearer to zero.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

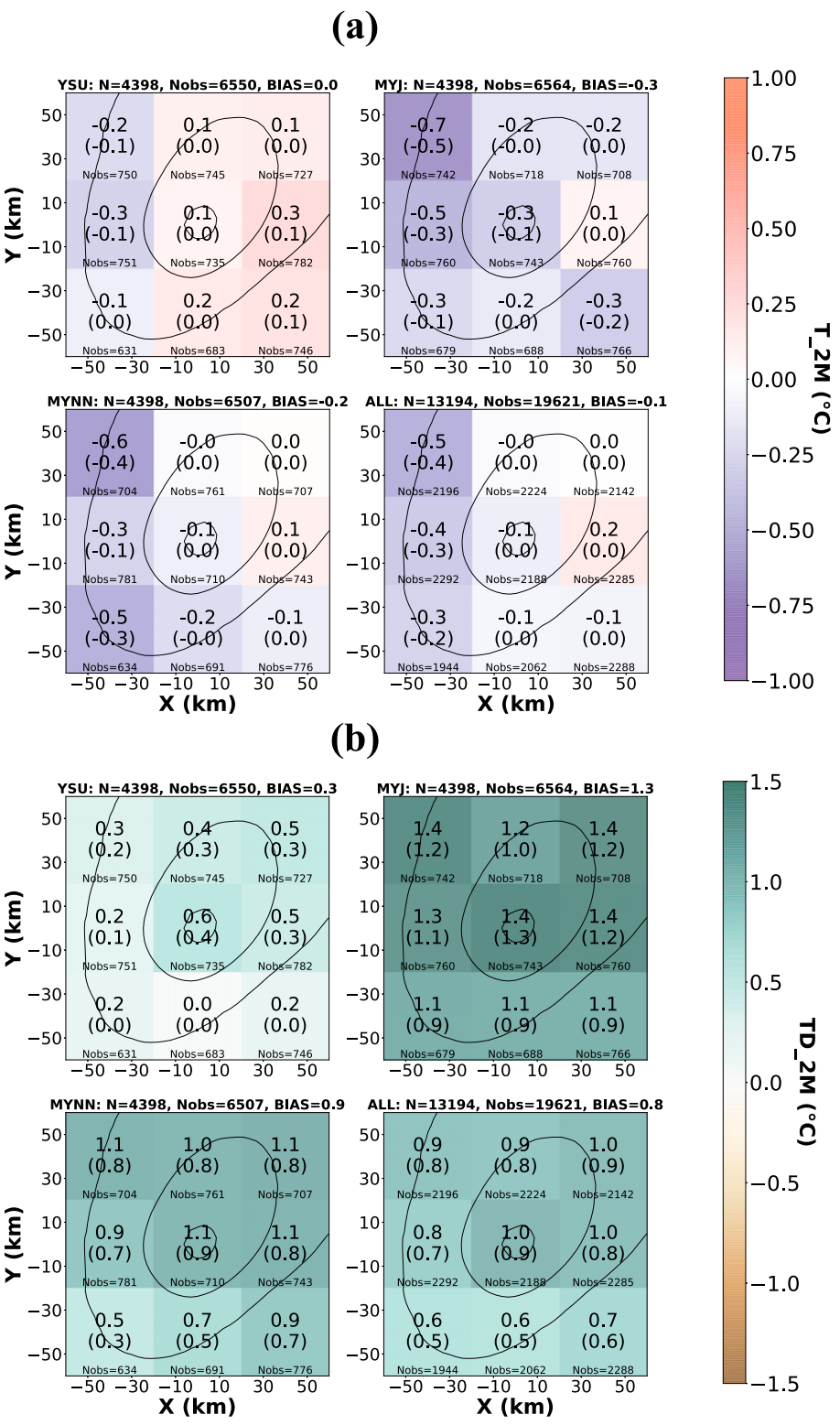

FIG. 3. Bias statistics for (a) T_2M and (b) TD_2M for each PBL subensemble and the full ensemble. The NSE-wide bias, number of analyzed storms N, and number of analyzed ASOS observations N_{obs} are listed at the top of each panel. The NSE is divided into nine subdomains that are shaded by their sample bias. Each subdomain's sample bias and confident lower bound on the bias (in parentheses) is annotated. The REFLCOMP probability-matched means are contoured (10, 30, and 50 dBZ) for reference.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

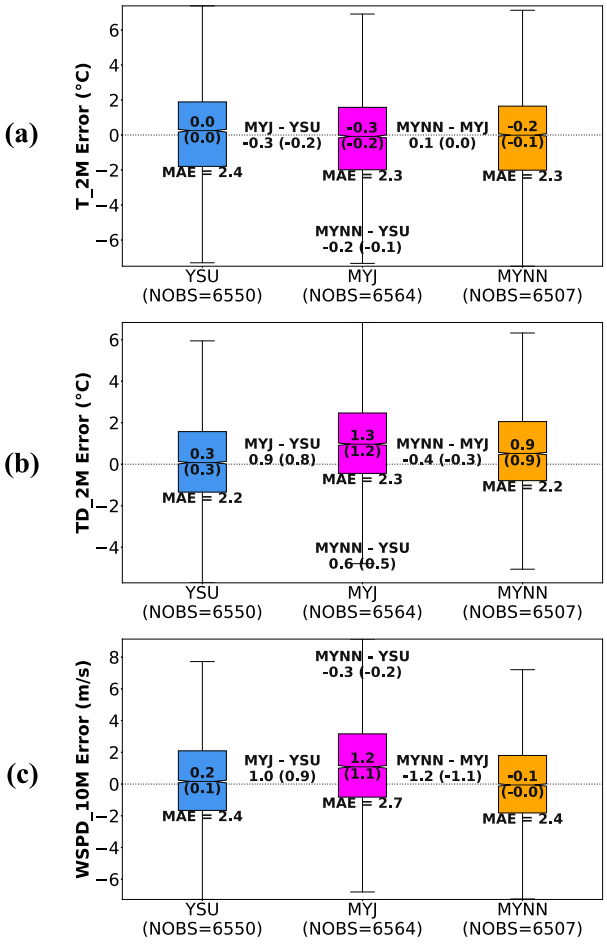

FIG. 4. Notched box-and-whisker plots of errors in (a) T_2M, (b) TD_2M, and (c) WSPD_10M for each subensemble. The boxes span the interquartile range (IQR), and the whiskers extend to half the IQR beyond the first and third quartiles. The notches span the 95% bootstrap confidence interval of the median ($N = 10\,000$). The medians and confident lower bounds (in parentheses) of the biases are annotated within the boxes; the medians and confident lower bounds (in parentheses) of the bias differences are annotated between the boxes. The median mean absolute errors (MAEs) are annotated below each box.

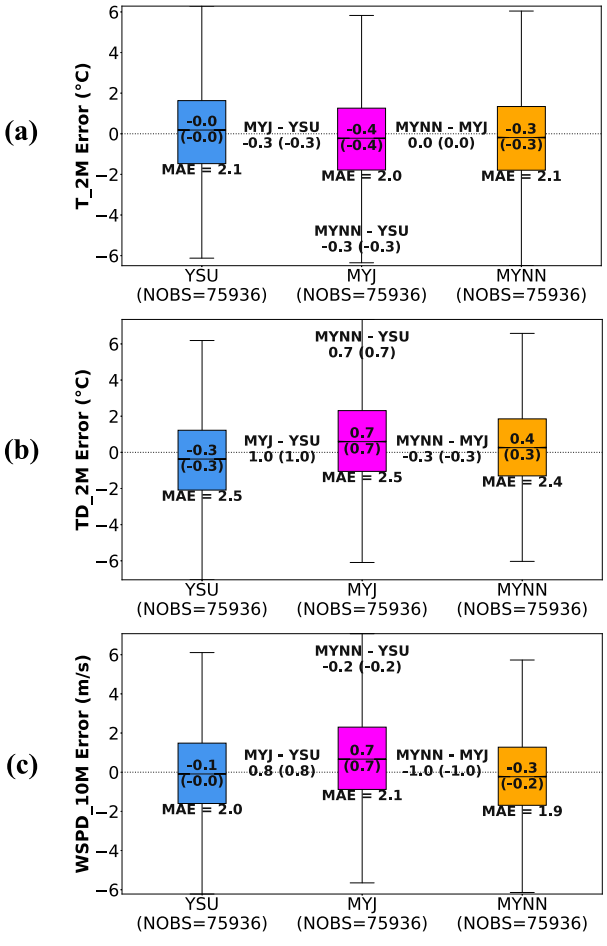

FIG. 5. As in Fig. 4, but using ASOS observations both within and outside NSEs.

consistent with previous studies that examined longer forecast lead times (Hu et al. 2010; García-Díez et al. 2013; Clark et al. 2015; Burlingame et al. 2017; Jahn and Gallus 2018). This result also comports with the deeper mixing often seen in nonlocal schemes, which account for countergradient fluxes (e.g., Cohen et al. 2015) and therefore tend to produce more entrainment at the PBL top (e.g., Hu et al. 2010). All three schemes were too moist on average, with the YSU scheme being the least biased with respect to both T_2M and TD_2M (biases of 0.0° and 0.3°C, respectively) and the MYJ scheme the most biased (biases of −0.3° and 1.3°C, respectively). In terms of WSPD_10M, the YSU and MYNN schemes were essentially unbiased, but the MYJ scheme had a positive bias of $1.2\,\mathrm{m\,s^{-1}}$ (Fig. 4c). Forecast biases and inter-PBL differences in U_10M and V_10M were $<0.5\,\mathrm{m\,s^{-1}}$ in magnitude (not shown). Recomputing the error distributions over all ASOS records within the WoFS domains (not just those within NSEs) revealed that interscheme differences in bias are similar both near and far from storms (Fig. 5). Thus, systematic interscheme differences in surface field magnitudes do not appear to be substantially enhanced or damped by simulated storm–environment interactions. This result is consistent with the spatial similarities between NSEs simulated with the different PBL schemes (Fig. 3).

To assess the signal-to-noise ratio of the PBL scheme impacts on surface variables in individual cases, we computed the mean standard deviation of each PBL scheme subensemble over the set of all ASOS observations, and then compared the standard deviations to the mean interscheme differences that were already presented in Fig. 5 (Table 2). The subensemble spread is generally at least as large as the mean differences between

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

TABLE 2. Mean PBL scheme subensemble standard deviations and mean inter-PBL-scheme differences valid across all ASOS observations.

	YSU stdev	MYJ stdev	MYNN stdev	MYJ–YSU	MYNN–MYJ	MYNN–YSU
T_2M	0.7	0.7	0.7	−0.3	0.0	−0.3
TD_2M	0.9	1.0	0.9	1.0	−0.3	0.7
WSPD_10M	0.9	1.0	0.8	0.8	−1.0	−0.2

subensembles, indicating that differences arising from systematic PBL scheme impacts are substantially masked by differences arising from other sources (e.g., different initial conditions) in individual cases. The similarity of the medians and confident lower bounds of the forecast biases and interscheme differences (Fig. 5), however, indicates that we sample a large enough number of cases to confidently isolate the systematic PBL scheme impacts despite the low signal-to-noise ratio in individual cases.

To explore how the forecast biases for T_2M, TD_2M, and WSPD_10M vary with the observed surface conditions, we perform two types of distribution-oriented verification (Murphy and Winkler 1987; Brooks and Doswell 1996). First, we examine forecast amplitude biases binned over prescribed intervals of observed values (Fig. 6). The amplitude biases for all three variables vary greatly with the observed conditions. In instances where rarer (i.e., nearer the tails of the frequency distribution) values of the variables are observed, the forecasts tend to have more common (i.e., nearer the mode of the frequency distribution) values, regardless of the PBL scheme. For example, the forecasts are, on average, much too warm in particularly cool conditions, and much too cool in particularly warm conditions (Fig. 6a). These forecast biases could result from values near both tails of the observed frequency distributions being predicted too rarely and, correspondingly, midrange values being predicted too often. In other words, the forecasts could exhibit narrower probability distributions than does the real atmosphere. There is another type of forecast error, however, that could contribute to the amplitude biases seen in Fig. 6. Even for forecast and observed probability distributions that are identical, unavoidable phase errors

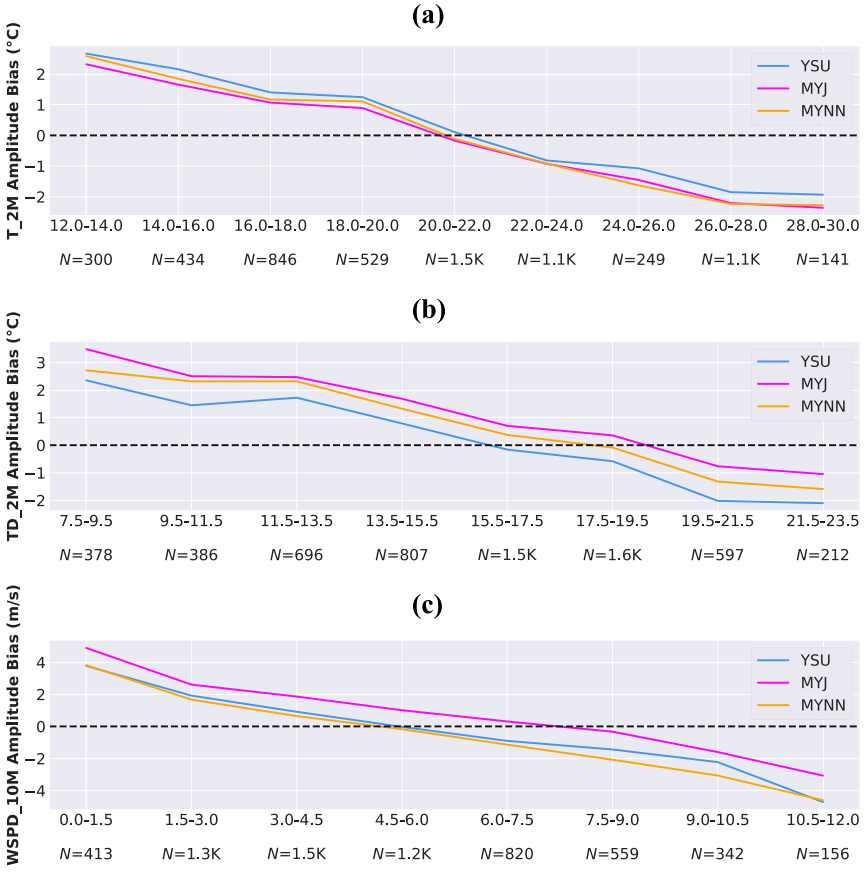

FIG. 6. Forecast amplitude bias valid for binned observed values of (a) T_2M, (b) TD_2M, and (c) WSPD_10M. Observation sample sizes are listed below corresponding bins.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

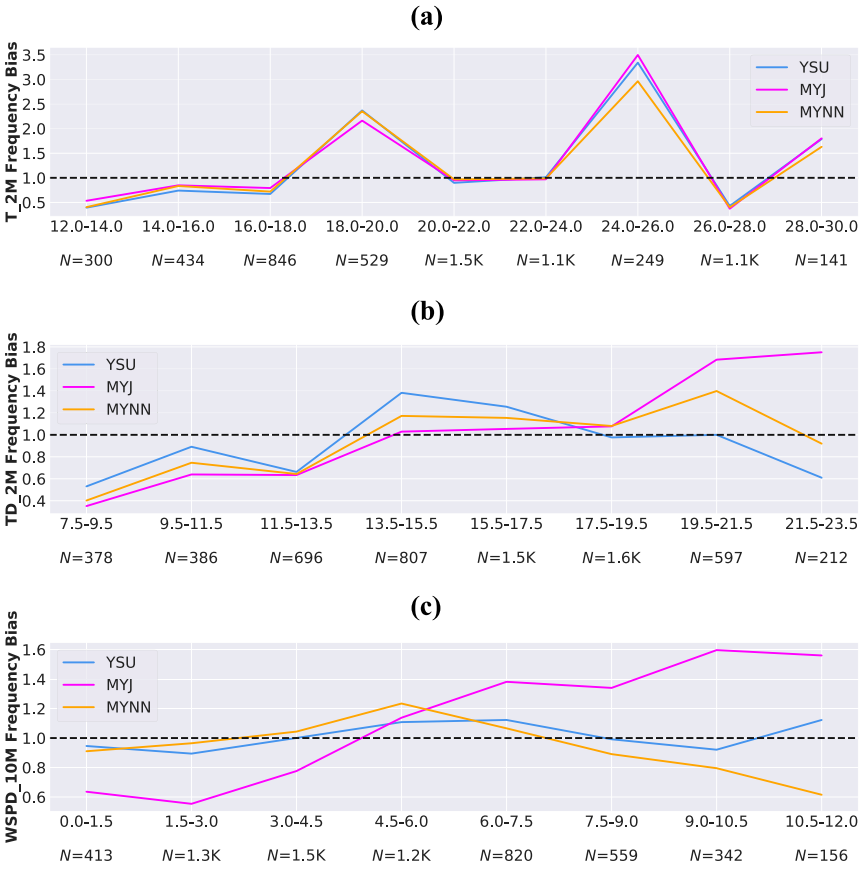

FIG. 7. As in Fig. 6, but for frequency bias.

in CAM forecasts will create a tendency for more-common values to be predicted in instances where rarer values are observed. This is because even a modest spatiotemporal offset between the forecast and observed fields in any given case would tend to cause rarer values in the observed field to overlap more-common values in the forecast field (and vice versa) by virtue of the larger coverage of the latter. To account for the effect of this sampling bias on the amplitude biases (Fig. 6), we additionally examine the *frequency* bias distributions (Fig. 7). This additional analysis indicates that much of the forecast amplitude bias in T_2M, TD_2M, and WSPD_10M arises from overly narrow forecast frequency distributions of these variables, and not solely from the sampling bias just discussed. For example, the too-moist forecasts in drier conditions and too-dry forecasts in moister conditions (Fig. 6b) can be explained largely by negative biases in the forecast frequency distributions of TD_2M near the tails of the observed distributions (Fig. 7b). Whether the frequency distribution biases arise primarily from limitations of the PBL schemes themselves or from some other model deficiency cannot be determined from this analysis alone; we briefly return to this question in section 6.

Not all of the amplitude biases in the surface variables can be explained by frequency distribution biases. For example, the MYJ forecasts produce too many higher-end WSPD_10M and too few lower-end WSPD_10M (Fig. 7c), which strongly suggests that the negative amplitude bias in higher-end WSPD_10M (Fig. 6c) arises from the aforementioned sampling bias, since the sampling bias can explain both the lower- and higher-end amplitude biases, whereas the frequency bias alone would produce a *positive* amplitude bias in higher-end WSPD_10M. Such detailed insights into the dependence of forecast bias on the observed atmospheric conditions could be exploited by calibration techniques to improve both deterministic and probabilistic forecasts. Distinguishing the effects of forecast probability distribution errors and phase errors (and resulting sampling bias) on amplitude biases may be important for optimizing forecast calibration, since the two types of error should ideally be treated differently.

The amplitude and frequency bias distributions associated with the different PBL schemes exhibit many similarities. For example, while the mean TD_2M

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

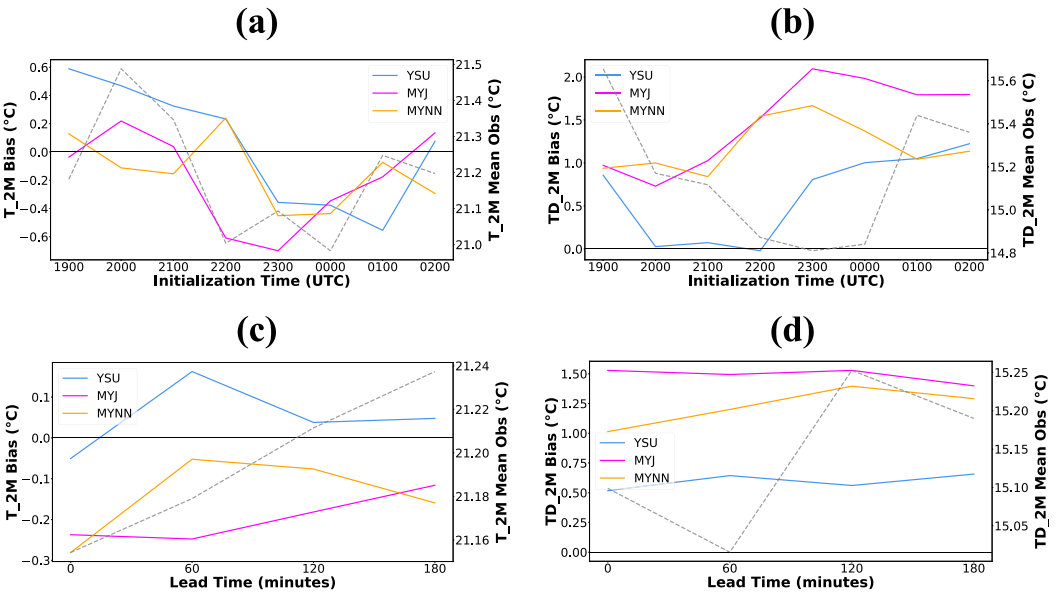

FIG. 8. Bias vs (top) initialization time and (bottom) lead time in (a),(c) T_2M and (b),(d) TD_2M. The dashed curves indicate the mean observed values of each variable for reference.

amplitude bias substantially differs among the schemes (as previously noted; e.g., Fig. 4), the interscheme difference in bias is relatively constant across the examined range of observed TD_2M (Fig. 6b). The frequency biases for all three variables differ more among the PBL schemes than do the amplitude biases (cf. Figs. 6, 7). Overall, however, the systematic impacts of the PBL schemes on surface forecast biases appear to be reasonably similar across a range of warm season conditions, which should make it easier to develop post hoc corrections for these biases. The insensitivity of interscheme *differences* in forecast amplitude biases combined with the much larger variations in the forecast amplitude biases *themselves* may largely explain why analyzed surface and PBL biases, and therefore assessments of the relative accuracy of different PBL schemes, qualitatively vary between prior studies (which often focus on different seasons, regions, and atmospheric scenarios from one another), despite analyzed error differences between the schemes generally agreeing across studies. For example, in our analysis, the predicted TD_2M average is approximately 1°C higher in the MYJ forecasts than in the YSU forecasts across the range of observed TD_2M (Fig. 6b); however, the YSU forecasts exhibit a much smaller TD_2M amplitude bias magnitude than the MYJ forecasts in lower-TD_2M conditions, but a much higher bias magnitude in higher-TD_2M conditions.

To assess how the PBL scheme impacts evolve with time, we examine time series of mean forecast bias in T_2M and TD_2M versus initialization time (aggregated over all lead times) and lead time (aggregated over all initialization times; Fig. 8). The relative differences between the YSU scheme and each other scheme (Figs. 4a,b) are valid at most of the initialization and all lead times, whereas the differences between the MYNN biases and the MYJ biases are more variable. Bias magnitudes in both variables neither steadily increase nor decrease with initialization and lead time, which suggests that while systematic PBL scheme errors do not rapidly accumulate as forecasts proceed, neither are

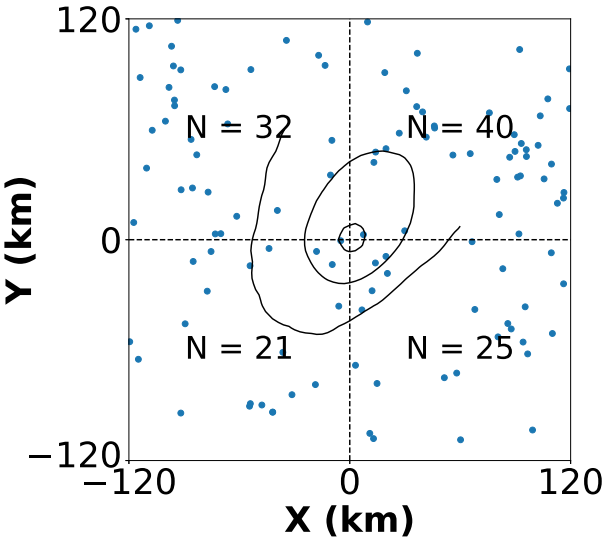

FIG. 9. Storm-relative release locations of soundings used to verify model near-storm vertical profiles.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

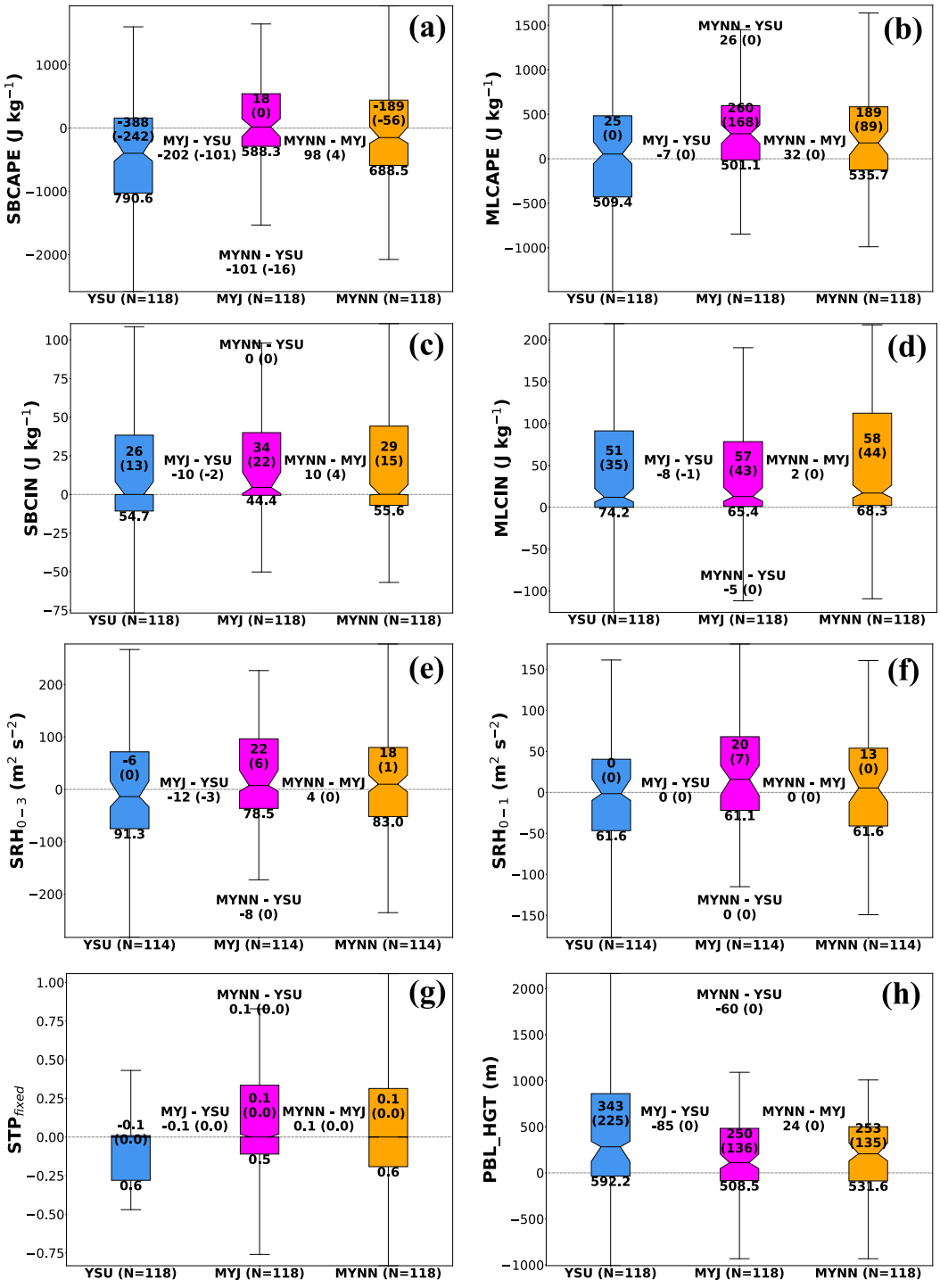

FIG. 10. Notched box-and-whisker plots of model errors in (a) SBCAPE, (b) MLCAPE, (c) SBCIN, (d) MLCIN, (e) $SRH_{0–3}$, (f) $SRH_{0–1}$, (g) STP_{fixed}, and (h) PBL_HGT. The medians and confident lower bounds (in parentheses) of the biases are annotated within the boxes with the medians and confident lower bounds (in parentheses) of the bias differences annotated between the boxes. The median MAEs are annotated below each box.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

they effectively damped by the data assimilation. The failure of a sophisticated, rapidly updating data assimilation system to satisfactorily mitigate surface biases is perhaps not surprising given the sparseness of surface and PBL observations, and the inability to efficiently assimilate observations in the presence of erroneous ensemble covariances. The T_2M biases in the analyses (0-min lead times in Fig. 8c) are slightly smaller (YSU) or slightly larger (MYJ, MYNN) than the biases in the NOAA Real-Time Mesoscale Analysis computed over the CONUS and for the same analysis times examined in the present study (Morris et al. 2020).

c. *Vertical profile verification*

Rawinsonde soundings valid within 240-km-diameter storm-centered domains (Fig. 9) are used to verify model vertical profiles of temperature, moisture, wind, pressure, and height, with emphasis on the lowest 1 km AGL. The model profiles are constructed at the WoFS grid point nearest the corresponding rawinsonde station. Sounding-derived parameters—MLCAPE, MLCIN, SBCAPE, SBCIN, MUCAPE, MUCIN, $SRH_{0–3}$, $SRH_{0–1}$, SCP, STP_{fixed}, and PBL_HGT (Table 1)—are computed from both the observed and model soundings, and then the model errors are compared across the PBL schemes (Fig. 10).

For many of the examined sounding parameters, the three PBL schemes produce very different forecast error distributions. For the majority of parameters, the YSU scheme produced the largest range of errors, and the MYJ scheme the smallest range of errors. In terms of SBCAPE, the MYNN and (especially) YSU schemes are negatively biased, while the MYJ scheme is essentially unbiased (Fig. 10a). The MYJ and MYNN schemes produce too much MLCAPE, while the YSU scheme is essentially unbiased in this parameter (Fig. 10b). All three schemes underpredict the magnitudes of SBCIN (Fig. 10c) and MLCIN (Fig. 10d). This could be related to the tendency for models to coarsely represent capping inversions (Coniglio et al. 2013). However, manual inspection revealed only 14% of the observed soundings contain capping inversions (this perhaps isn't surprising since we used only those soundings collected near ongoing convection), which suggests another, unknown factor is contributing to the underprediction of convective inhibition. YSU forecasts are essentially unbiased in $SRH_{0–3}$ and $SRH_{0–1}$, while MYJ and MYNN are weakly positively biased (Figs. 10e,f). Cohen et al. (2017), who analyzed cold-season environments over the Southeast United States, likewise found that YSU produces lower SRH than MYJ. YSU produces lower STP_{fixed} than the other two schemes (Fig. 10g), which is not surprising considering the interscheme differences in MLCAPE and $SRH_{0–1}$ (Figs. 10b,f), two of the four variables in the STP_{fixed} calculation.

TABLE 3. As in Table 2, but for all soundings launched within WoFS domains.

	YSU stdev	MYJ stdev	MYNN stdev	MYJ–YSU	MYNN–MYJ	MYNN–YSU
SBCAPE	257	306	280	−39	−38	−76
MLCAPE	196	243	221	44	−23	21
SBCIN	36	36	35	−3	3	0
MLCIN	33	34	33	−19	11	−7
$SRH_{0–3}$	36	36	36	−6	0	−6
$SRH_{0–1}$	24	26	25	3	0	2
STP_{fixed}	0.2	0.3	0.2	0	0	0
PBL_HGT	364	326	320	−175	38	−138

Computing PBL heights (PBL_HGT) from the virtual potential temperature profile as in Coniglio et al. (2013), we find that all three schemes produce PBLs that are too deep, with YSU producing the largest PBL_HGT overestimates (Fig. 10h), consistent with the scheme's tendency to overmix the PBL. While previous studies have also found that the YSU scheme tends to produce larger PBL heights than other schemes, the same studies concluded that MYJ underestimates, not overestimates, PBL heights (Hu et al. 2010; García-Díez et al. 2013; Coniglio et al. 2013). It should be noted that the methods used to compute PBL heights varied among these studies, as did factors important to simulated PBL height growth rates, including grid spacing and the relative representation of buoyant versus mechanical turbulence production regimes (Deardorff 1972; Moeng and Sullivan 1994). Equally noteworthy, however, is the similarity of Coniglio et al. (2013) to the present study, but for our focus on much shorter forecast lead times than in other studies. The question of whether this tendency for even local schemes to overdeepen the PBL is particular to the HRRRE and/or WoFS configurations used in the present study or generally obtains at $O(1)$ h lead times would be worth pursuing in future work.

In many instances, the confidence intervals on the sounding parameter biases and interscheme bias differences are very large, as indicated by large differences between the medians and confident lower bounds in Fig. 10. This large uncertainty in the sounding parameter biases and bias differences contrasts with the highly confident surface variable error analyses (Figs. 4, 5). This can be explained in part by the much smaller number of soundings than ASOS observations, but repeating the signal-to-noise ratio analysis that was performed for all surface observations (Table 2) for all soundings collected within WoFS domains (i.e., whether near a storm or not) reveals that systematic PBL scheme impacts are dominated by intra-sub-ensemble differences in individual cases (Table 3). These results motivate averaging over all subensemble members containing a

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

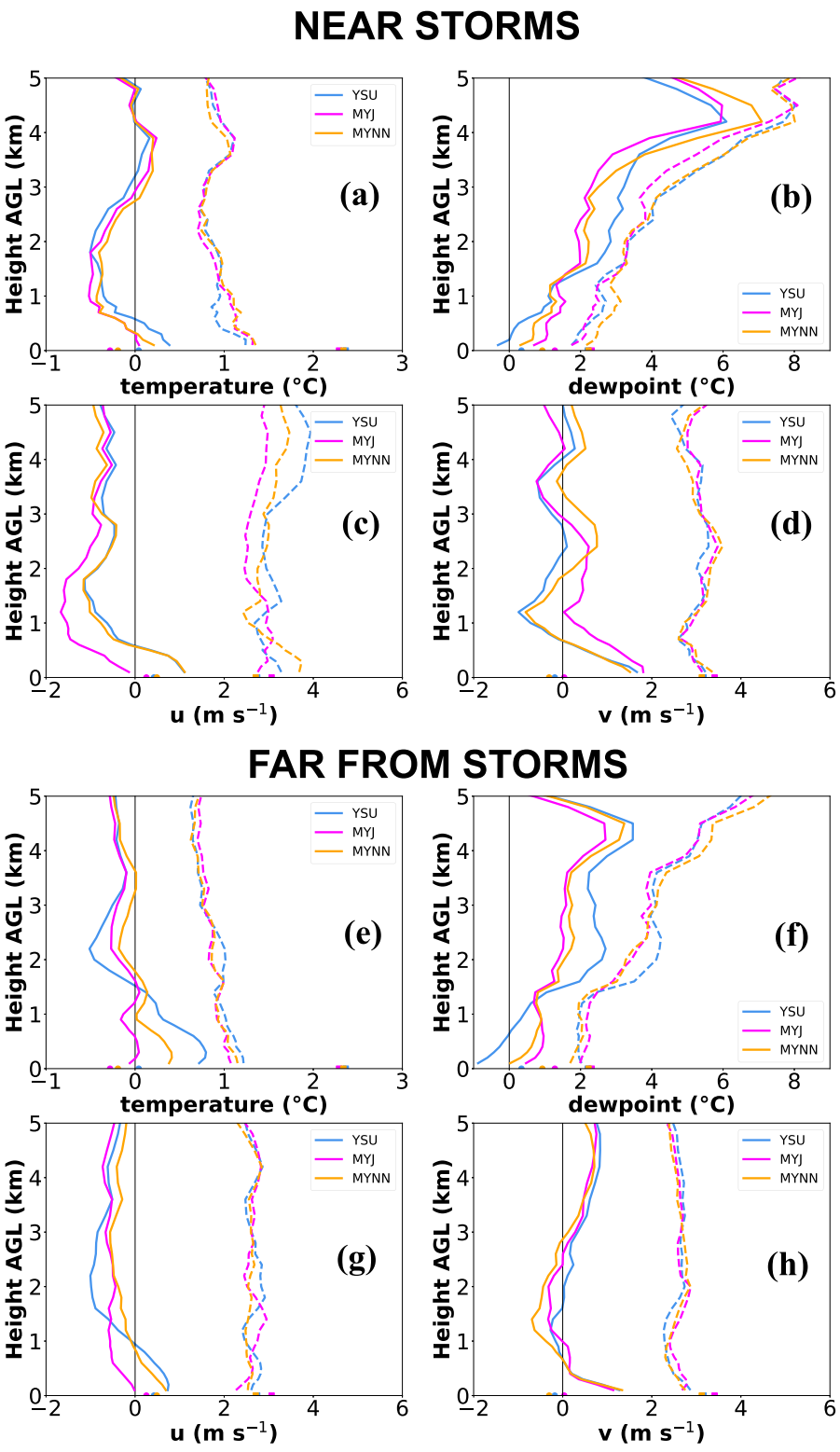

FIG. 11. (a)–(d) Vertical profiles of model bias (solid) and MAE (dashed) in (a) temperature, (b) mixing ratio, (c) u, and (d) v within the 240-km NSEs ($N = 118$). Surface biases (semicircles) and MAE (rectangles) in (a) T_2M, (b) TD_2M, (c) U_10M, and (d) V_10M were computed from ASOS observations collected within the NSEs. (e)–(h) As in (a)–(d), but using soundings outside the 240-km NSEs ($N = 571$) and all ASOS observations collected within the WoFS forecast domains.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

matched storm (or using a much larger number of cases) in future work in order to obtain more confident estimates of biases and interscheme differences.

Next, vertical profiles of model bias and MAE (Fig. 11) are computed for each PBL scheme by interpolating the observed and model soundings to a common vertical grid. The vertical grid begins at 100 m AGL and proceeds in 100-m increments to 1000 m AGL, 200-m increments from 1200 to 3000 m, and 300-m increments from 3300 to 5100 m. The inter-PBL differences within the lowest 1 km of the vertical profiles of bias (Figs. 11a–d) are qualitatively consistent with the inter-PBL differences in ASOS biases. For example, the YSU and MYJ profiles are the warmest/driest and coolest/moistest, respectively (Figs. 11a,b), consistent with many previous studies (Hu et al. 2010; García-Díez et al. 2013; Clark et al. 2015; Kerr et al. 2017; Burlingame et al. 2017; Jahn and Gallus 2018). The warmer, drier YSU profiles contribute to the lower YSU forecast values of SBCAPE, MUCAPE, and STP_{fixed} noted above. All three schemes exaggerate PBL lapse rates (Fig. 11a). YSU produces the smallest temperature MAE in the lowest 1 km AGL (Fig. 11a), while MYNN produces both the largest temperature MAE and (by far) dewpoint MAE (Figs. 11a,b), consistent with Coniglio et al. (2013). No one scheme performs categorically better than the other with regard to low-level winds (Figs. 11c,d), though large interscheme differences in error characteristics occur at individual levels.

Repeating the analysis for all soundings collected within the WoFS domains (Figs. 11e–h) and comparing to the near-storm analysis suggests that the more prominent interscheme PBL profile differences are similar whether near or far from storms. Some of the differences between the near-storm and all-soundings analyses are likely due to sampling errors, especially in the former ($N = 118$), making it difficult to attribute small differences to changes in PBL scheme behavior nearer versus farther from storms. Together with the previously demonstrated similarity of interscheme differences in surface biases near versus far from storms (cf. Figs. 4, 5), these results suggest that storms do not strongly modulate model differences arising between the different PBL schemes.

Surface bias and MAE are computed for each variable in Fig. 11 using ASOS observations collected within NSEs (Figs. 11a–d) or throughout the WoFS forecast domains (Figs. 11e–h). While the ordering of the surface biases for the different PBL schemes generally reflects the ordering of the biases over the lowest 1-km AGL, substantial vertical discontinuities appear between the surface errors and 100-m AGL errors in most instances. We speculate that these discontinuities arise in part from the diagnostic nature of the surface variables, which are largely determined by the land surface model and surface layer schemes and therefore only indirectly linked to the prognostic variables on the model vertical grid. Differences in the scales represented by the observed and model variables, and how these scale mismatches themselves differ between variables within versus just above the surface layer, may also contribute to the vertical discontinuities in errors.

	YSU	MYJ	MYNN
T_2M (°C)	0.0	-0.3	-0.2
TD_2M (°C)	0.3	1.3	0.9
WSPD_10M (m s^{-1})	0.2	1.2	-0.1
SBCAPE (J kg^{-1})	-388	19	-190
MLCAPE (J kg^{-1})	25	260	189
SBCIN (J kg^{-1})	26	34	29
MLCIN (J kg^{-1})	51	57	58
$SRH_{0\text{-}3}$ (m^2 s^{-2})	-6	22	18
$SRH_{0\text{-}1}$ (m^2 s^{-2})	0	19	13
STP_{fixed}	-0.1	0.1	0.1
PBL_HGT (m)	343	250	253

FIG. 12. PBL scheme forecast biases computed from near-storm ASOS and sounding verification. Green, yellow, and red shading indicate subjectively determined small, medium, and large biases, respectively, in forecast variables.

The biases in each of the model surface and sounding-derived variables are listed for each PBL scheme in Fig. 12. While no scheme performed categorically better than the others with respect to these metrics, YSU produced forecast biases that were similar to or smaller than those produced by the MYJ and MYNN schemes for 9 of the 11 variables. Of course, there are numerous other forecast metrics that could be considered, and the optimal choice of PBL scheme will likely depend upon the application.

5. Physics impacts on storm and near-storm environment (nse) characteristics

a. *Probability-matched means of storm and NSE fields*

Probability-matched means (PMMs; Ebert 2001) of NSE fields are computed as in P19. PMMs preserve the average probability distribution of constituent cases, thus mitigating the damping of extrema and

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

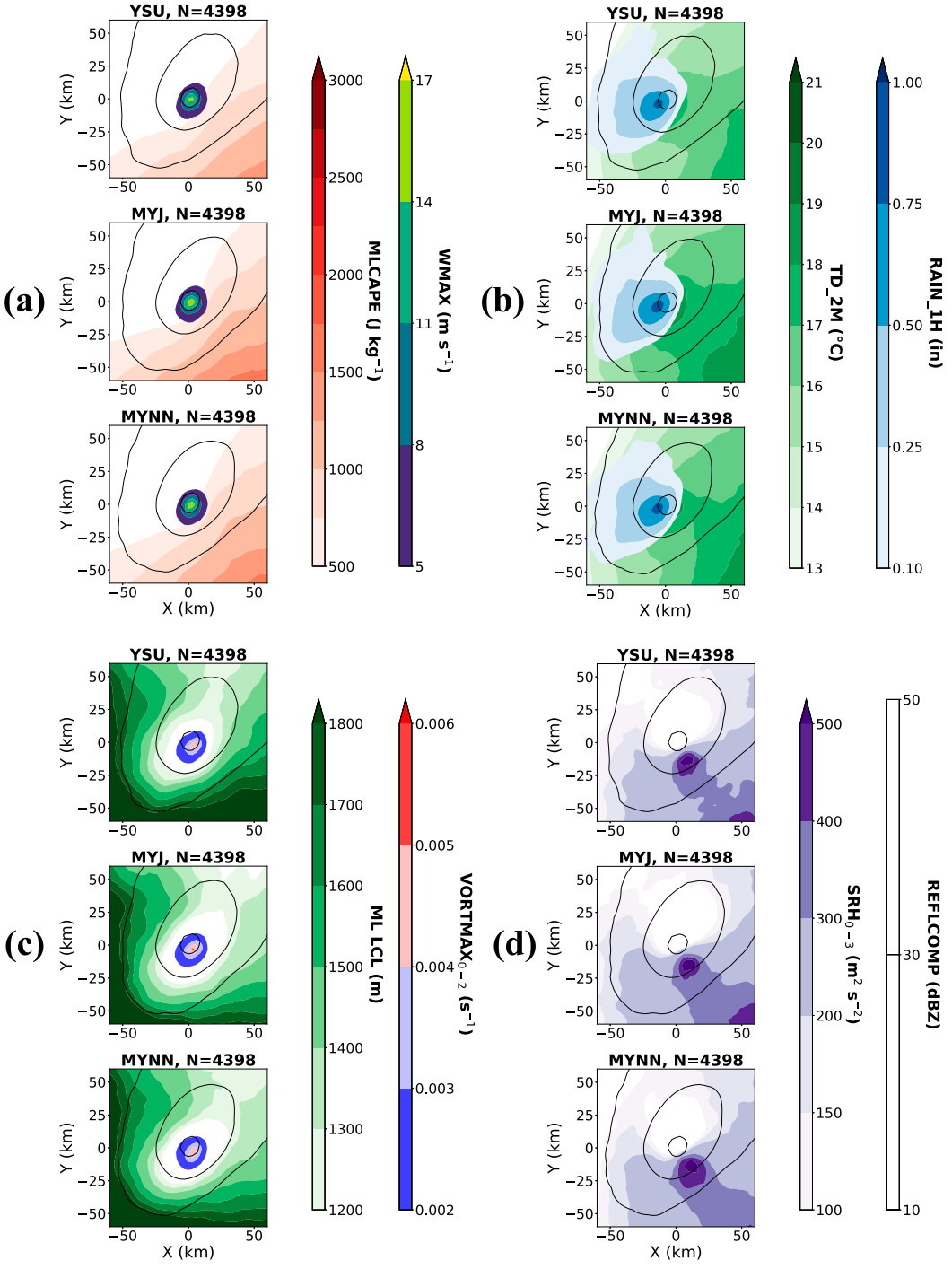

FIG. 13. Probability-matched mean (a) MLCAPE and WMAX, (b) TD_2M and RAIN_1H, (c) MLLCL and MAXVORT, and (d) SRH_{0-3}. REFLCOMP is contoured at 10, 30, and 50 dBZ for reference.

gradients that occurs with simple averaging. The spatial gradients in the PMM fields vary little across the PBL schemes (Fig. 13). Differences in the field magnitudes are consistent with the error differences noted in section 4; for example, the YSU PMMs exhibit lower MLCAPE (Fig. 13a) and TD_2M (Fig. 13b) than the other schemes.

One of the more interesting features visible in the PMMs is worth a brief digression. The SRH_{0-3} shows a maximum about 20 km southeast of the storm in the

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

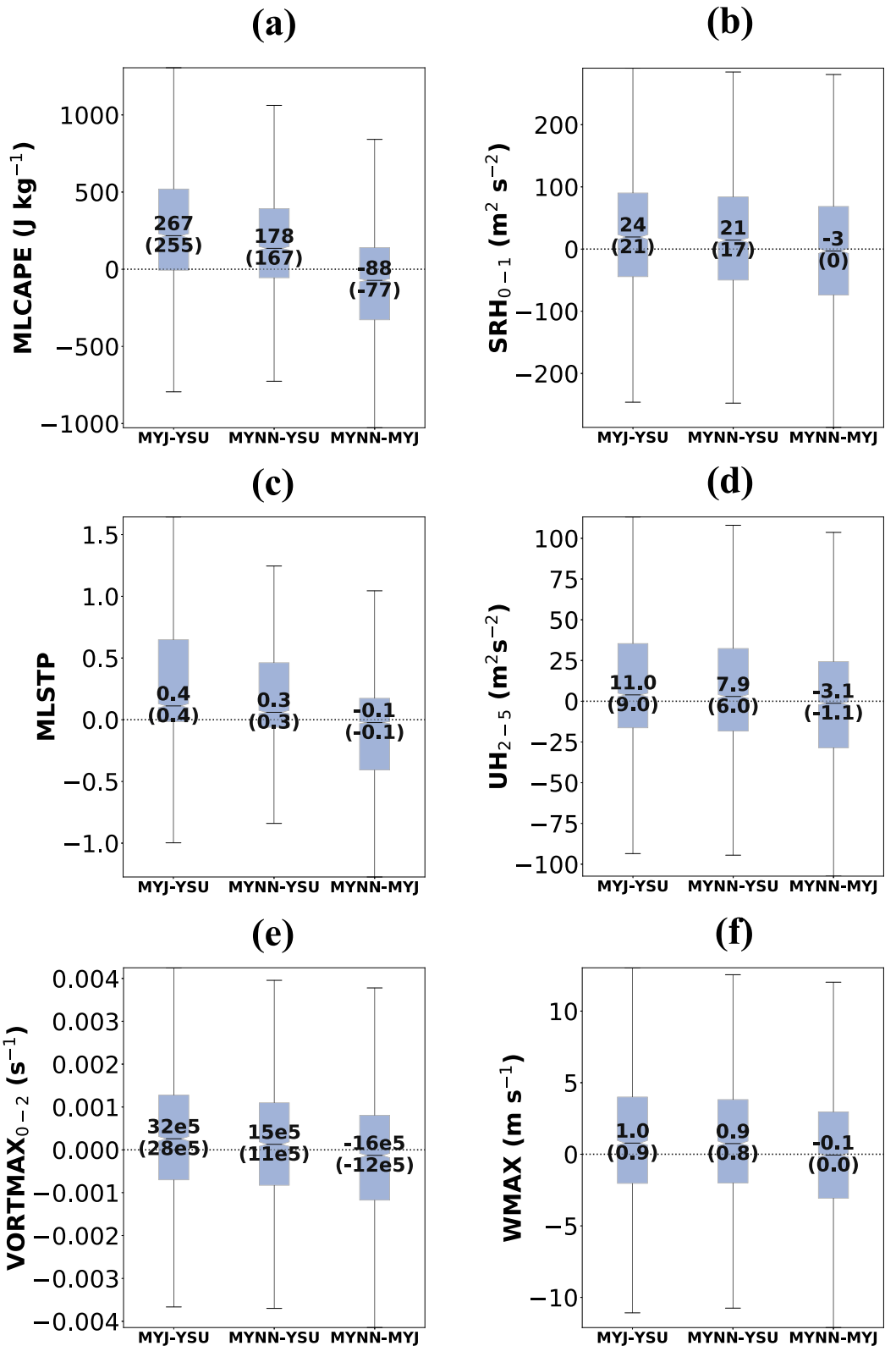

FIG. 14. Notched box-and-whisker plots of differences in NSE maxima between each pair of PBL schemes: (a) MLCAPE, (b) $SRH_{0–1}$, (c) MLSTP, (d) $UH_{2–5}$, (e) VORTMAX, and (f) WMAX.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

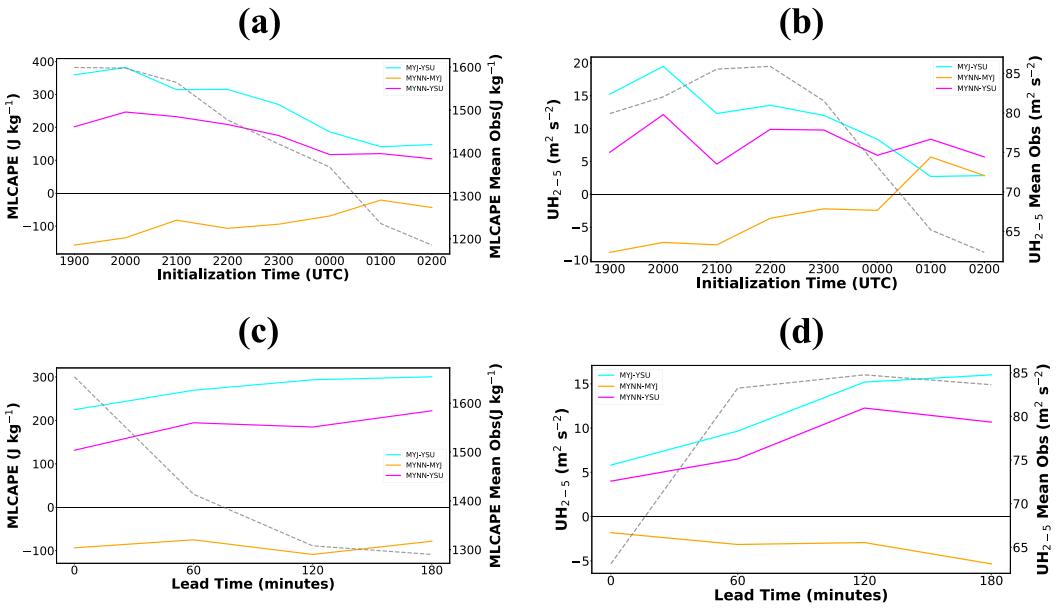

FIG. 15. Interscheme differences in NSE maxima vs (top) initialization time and (bottom) lead time: (a),(c) MLCAPE, and (b),(d) UH_{2-5}. The dashed curves indicate the mean absolute values of each variable for reference.

region where supercells are known to enhance the low-level wind shear in the near-field environment (Parker 2014; Wade et al. 2018). This feature, which also arises in next-day forecasts (P19), is evidence that 3-km grid spacing can at least qualitatively capture the near- and in-storm perturbation low pressure associated with the storm–environment interactions that are hypothesized to be the primary driver of the enhanced low-level inflow and shear. These results support the hypothesis that 3-km grid spacing is sufficient to realistically simulate many of the important processes within supercells (e.g., Potvin and Flora 2015).

b. *Distributions of NSE maxima*

NSE-wide (i.e., within each storm-centered 120-km box) maxima[6] are computed for several model fields for each PBL scheme. Differences are then computed between each pair of schemes (Fig. 14). The MYJ and MYNN schemes, on average, produce larger MLCAPE (Fig. 14a), SRH_{0-1} (Fig. 14b), and STP_{fixed} (Fig. 14c) than the YSU scheme, consistent with the sounding parameter verification (Fig. 10). This translates to generally more strongly rotating storms (Figs. 14d,e) with stronger updrafts (Fig. 14f) and attendant hail sizes (not shown) with the MYJ and MYNN schemes. The MYJ–MYNN differences in NSE maxima are much smaller than the MYJ–YSU and MYNN–YSU differences, with the MYNN scheme on average producing slightly less favorable NSEs and weaker storms than the MYJ scheme. The relative similarity between the MYJ and MYNN results is perhaps not surprising given that both schemes are based on the Mellor–Yamada 1.5-order local scheme (Mellor and Yamada 1974, 1982).

To assess how the PBL scheme impacts evolve with time, we examine time series of mean interscheme forecast differences versus initialization time (aggregated over all lead times) and lead time (aggregated over all initialization times; Fig. 15). Interscheme differences in both the storm environment (e.g., Fig. 15a) and storm attributes (e.g., Fig. 15b) generally decrease with initialization time, possibly due in part to the decreases in MLCAPE and UH_{2-5} themselves over this period (Figs. 15a,b). However, the interscheme differences either increase with lead time (Figs. 15c,d) or exhibit very little trend overall.

c. *Distributions of storm attributes*

Inter-scheme differences in several storm attributes were examined using a procedure similar to that used for the NSE maxima intercomparisons. As shown in Table 4, the mean area, length, and orientation of storms did not meaningfully vary among the PBL schemes, nor

TABLE 4. Mean characteristics of observed (MRMS) storms and storms simulated using the YSU, MYJ, and MYNN schemes.

	YSU	MYJ	MYNN	MRMS
Storm area (km^2)	445	461	450	340
Storm length (km)	34.4	34.8	34.0	33.3
Storm orientation (°)	−10	−10	−11	−12
Cold pool area (km^2)	644	686	703	N/A
Updraft area (km^2)	150	144	140	N/A

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

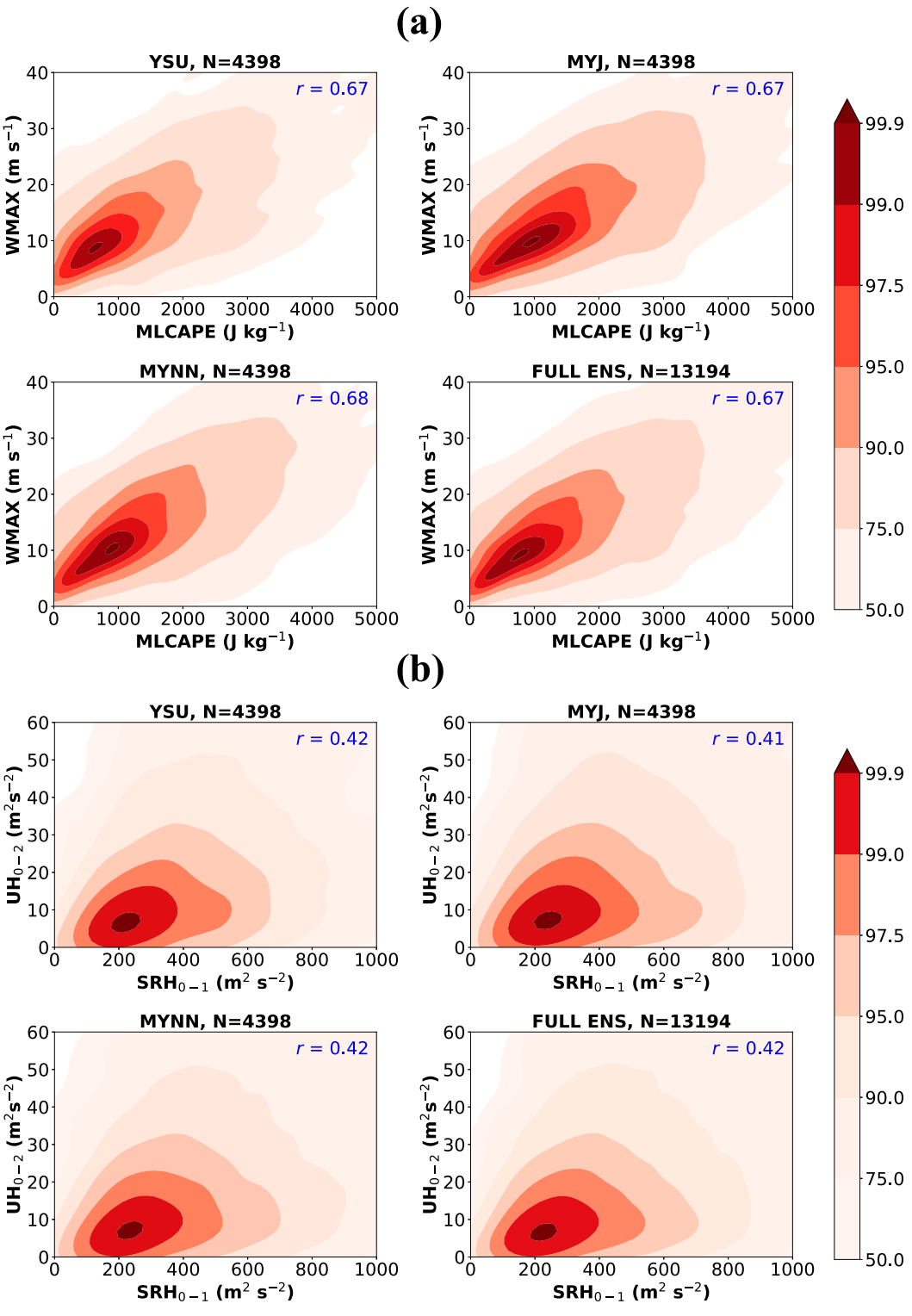

FIG. 16. Kernel density estimates of (a) MLCAPE and WMAX, and (b) $SRH_{0–1}$ and $UH_{0–2}$. The Pearson correlation coefficient r is shown on each panel.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

did the mean area of storm updrafts (identified using a w threshold of $2.5\,m\,s^{-1}$) or cold pools (identified using a -1.7°C threshold for deviation of T_2M from the NSE mean). Thus, while systematic NSE differences between schemes produced nontrivial differences in storm intensity (section 5c), the NSE differences did not substantially impact storm morphology.

d. *Bivariate kernel density estimates*

As in P19, bivariate kernel density estimates (KDEs) are examined for selected pairs of statistics to identify systematic differences in the relationships between storm and/or NSE characteristics (Fig. 16). KDEs were computed for: MLCAPE and WMAX (Fig. 16a), $SRH_{0–1}$ and $UH_{0–2}$ (Fig. 16b), MLCAPE and MAXHAIL, MLCAPE and $UH_{2–5}$, MLCAPE and $SRH_{0–3}$, and $CAPE_{0–3}$ and $UH_{0–2}$. No substantial interscheme differences arose in any of the bivariate relationships examined, suggesting the choice of PBL scheme did not fundamentally alter important storm–environment interactions.

e. *Daily mean analyses*

To assess the variability of PBL impacts across different convective events, we repeat some of our analyses on each daily set of forecasts containing at least 30 storm objects ($N = 36$). Computing PMMs (section 5a) for individual days (not shown) reveals that differences between the geometry of NSEs simulated with different PBL schemes vary nearly as little on individual days as over the full 40-day dataset. Thus, even on a regional scale, the choice of PBL scheme does not appear to substantially impact the qualitative spatial patterns of NSEs at 0–3-h lead times. Daily analyses of field magnitudes likewise confirm the representativeness of the all-days results. Computing surface variable biases (section 4b) for individual days (Fig. 17) noticeably reduces interscheme overlap in bias distributions relative to the all-days biases (cf. Figs. 4, 17). Similarly, interscheme differences in NSE maxima (section 5b) valid for individual days (Fig. 18) are larger than for the full dataset (cf. Figs. 14, 18). These two results indicate the relative PBL scheme impacts documented in previous sections are not confined to a narrow range of atmospheric scenarios, but instead frequently recur from day to day during the warm season. The repeatability of the interscheme forecast differences, which could also be inferred from the amplitude bias distributions (Fig. 6), suggests that it would be straightforward to develop postprocessing techniques to mitigate forecast biases associated with different PBL schemes.

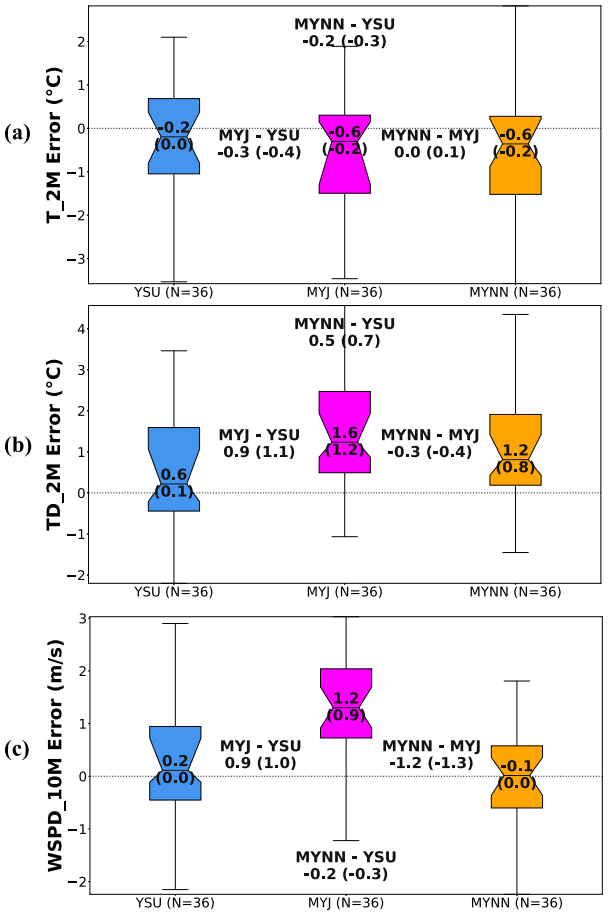

FIG. 17. As in Fig. 4, but for daily biases.

6. Summary and discussion

The systematic impacts of the three PBL schemes used in the NSSL Warn-on-Forecast System (WoFS) were assessed using a novel evaluation framework tailored to thunderstorms and near-storm environments (NSEs). This storm-based framework is being developed in part to improve our understanding of the impacts of CAM design choices on the simulation and prediction of thunderstorm initiation, track, and intensity. Such knowledge will be crucial for improving operational forecasts of thunderstorm hazards at lead times of minutes to days.

Very short forecast lead times of 0–3 h were examined in the present study, which facilitated comparisons between observational analyses and different model simulations of storms and thereby increased the precision with which PBL scheme impacts could be measured. ASOS verification revealed that all three WoFS PBL schemes–YSU, MYJ, and MYNN–exhibit a cool, moist bias at the surface, with the YSU members producing the least bias, consistent with the well-documented larger mixing and associated warming and drying associated with this and other nonlocal PBL schemes relative to the MYNN and (especially) MYJ schemes. Rawinsonde verification revealed that the qualitative interscheme differences in surface temperature and

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

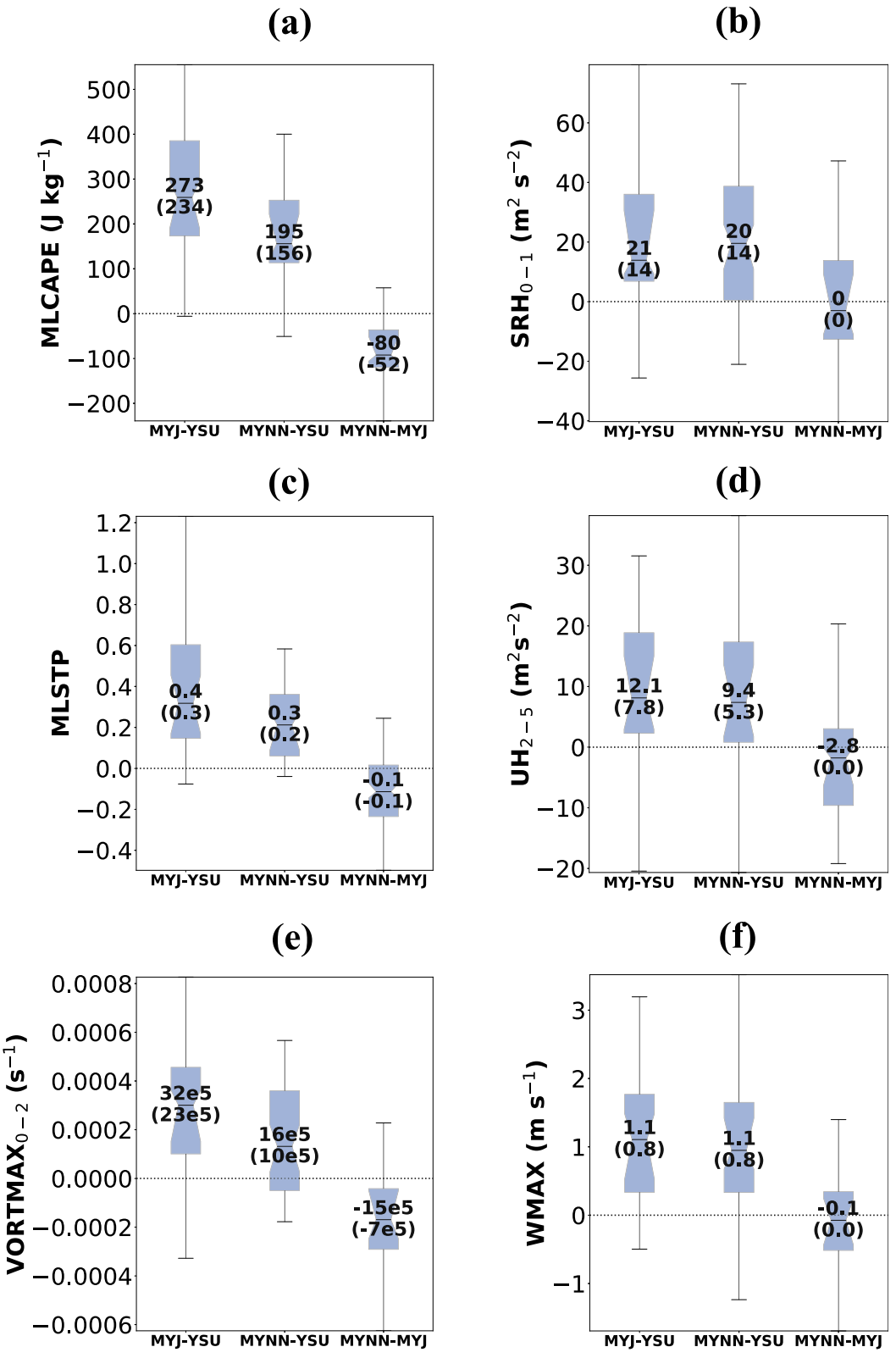

FIG. 18. As in Fig. 14, but for daily biases.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

dewpoint extended through at least 1 km AGL. YSU produced the least biased estimates of many important environmental variables, but substantially underestimated surface-based CAPE. All three schemes substantially underpredicted convective inhibition, which may in part explain the storm frequency biases identified in the WoFS (Skinner et al. 2018) and other ARW CAM ensembles (Potvin et al. 2019).

Forecasts using all three schemes produced too narrow a distribution of surface temperature, dewpoint, and wind speed. There are a number of potential explanations for this model deficiency. For example, PBL schemes were not originally designed to parameterize the finescale processes that CAMs would ideally capture; if PBL scheme limitations do indeed contribute to the overly narrow surface forecast distributions, then perhaps scale-aware/scale-adaptive PBL schemes (e.g., Olson et al. 2019) will be capable of accurately representing a wider range of atmospheric conditions. It is also possible that current land surface models are incapable of accurately parameterizing the full spectrum of warm-season atmospheric conditions. Insufficient spread in land surface conditions could also have contributed to the problem; the WoFS does not perturb the land surface variables, and the HRRRE in 2017–18 only perturbed soil moisture, and only at the start of cycling each day. The same overly narrow surface variable distributions are also present in the HRRRE forecasts used to initialize the WoFS (not shown), and therefore cannot be a consequence of rapid radar and/or satellite data assimilation or some other special aspect of the WoFS. Finally, the same model bias is present in regions distant from storms (not shown), and therefore cannot be attributed to storm–environment interactions.

While surface biases varied greatly with the observed conditions, the systematic interscheme differences in surface forecasts were much less sensitive. Surface and PBL biases were similar near versus far from storms, indicating that systematic differences in PBL scheme impacts are not strongly modulated by storm–environment interactions. Surface bias magnitudes did not generally increase with forecast lead time, suggesting that PBL scheme errors do not rapidly accumulate as forecasts proceed. On the other hand, the biases were not generally lower for later forecast initialization times, suggesting that systematic PBL scheme errors are not effectively damped by data assimilation, presumably due in part to the sparseness of surface and PBL observations.

The choice of PBL scheme did not systematically change the spatial configuration of NSE fields, but did impact the magnitudes of both environmental and storm intensity parameters. The YSU ensemble members produced weaker storms than the MYJ and MYNN members, which is likely attributable at least in part to the less favorable storm environments simulated in the YSU members. Examination of relationships between storm intensity and environmental parameters suggested the choice of PBL scheme did not fundamentally alter important storm–environment interactions. This makes it more probable that the storm intensity differences among the three PBL schemes arise primarily from environmental differences. Despite substantially modifying storm environments and storm intensity, the choice of PBL scheme did not substantially impact storm morphology or the skill of storm location predictions. Thus, while it may be worthwhile to weight WoFS ensemble member forecasts differently based on their PBL scheme, especially given the repeatability of the interscheme forecast impacts across different events, this approach is only expected to modestly improve probabilistic forecast guidance for convective hazards.

We plan to extend the analysis methods developed herein to assist in evaluating potential WoFS configuration changes, including increased horizontal and vertical model resolution, implementation of scale-aware PBL schemes, and transition to the FV3 dynamic core. The analysis methods could also be used to compare storm and NSE characteristics (both observed and WoFS-simulated) in different scenarios, for example, in Plains-type versus low-CAPE/high-shear regimes or on days with high versus low WoFS forecast skill. The present framework could be modified to assess PBL scheme impacts on mesoscale convective systems and their environments. Finally, this study was primarily concerned with PBL scheme impacts on simulations of existing storms; it would be valuable to extend this analysis to the prestorm environment and convection initiation.

Acknowledgments. This work was prepared by the authors with funding from the NSSL Forecast Research and Development Division (CKP, MCC, AJC, ENS) and the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce (PSS, KAH, JAG, MLF, AER). We thank Derek Stratman for helpful discussions and Christopher Kerr for informally reviewing an earlier version of this paper. We especially thank the two anonymous reviewers whose comments greatly enhanced the final manuscript. Valuable local computing assistance was provided by Gerry Creager, Karen Cooper, and Jeff Horn. All analyses and visualizations were produced using the freely provided Anaconda Python distribution. The contents of this paper do not necessarily reflect the views or official position of any organization of the United States.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

APPENDIX

Storm Tetrad Considerations

Herein we present three cautionary notes about our storm tetrad approach to assessing systematic impacts of different PBL schemes. First, the correspondence between the storms comprising each tetrad should not be interpreted too literally. Forecast and observed storms originating at substantially different locations or times can end up in the same area at the same time by happenstance; for example, during the forecast, a spurious model storm may initiate near a real, mature storm that developed well upstream. Observed and model storms can also evolve very differently (e.g., exhibit different convective modes) even when they do develop and move in proximity to one another. Our method therefore does not ensure that all three model storms in a tetrad closely represent the observed storm. This limitation likely reduces the signal-to-noise ratio of the systematic PBL scheme impacts, and could be addressed using object matching methods that enforce similarity in storm morphology and intensity (e.g., Johnson et al. 2020). Still, by comparing only model storms that occurred in proximity to an observed storm and therefore to each other, we ensure that comparisons between storms that developed in disparate environments are quite rare.

Second, in an ensemble with systematic differences in initial conditions (ICs) between members using the same PBL scheme, our method of selecting the first storm for each PBL scheme found in proximity to a given observed storm would be inappropriate. It would instead be necessary to use an approach that selects model storms randomly among the members of each PBL subensemble to avoid analyzing a combination of systematic interscheme differences and systematic IC differences. To account for the potential existence of unexpected systematic IC differences among the WoFS members, we repeated many of the analyses shown in this paper using storm tetrads created from same-physics (both the PBL and radiation schemes) ensemble members. The mean differences between same-physics storms were generally statistically insignificant, indicating that the systematic inter-PBL-scheme differences obtained in this study are not substantially modified (if at all) by any systematic intermember IC differences that unexpectedly occur in the WoFS.

Third, rather than selecting a single matched storm (and NSE) per PBL scheme for each intercomparison (i.e., our storm tetrad approach), we could have averaged over all subensemble members containing a matched storm. This would likely have increased the signal-to-noise ratio of the systematic PBL scheme impacts in our analyses by damping the influence of initial condition differences. Making this modification could substantially reduce the rate of type-II errors in future applications of our methodological framework.

REFERENCES

Adams-Selin, R. D., A. J. Clark, C. J. Melick, S. R. Dembek, I. L. Jirak, and C. L. Ziegler, 2019: Evolution of WRF-HAILCAST during the 2014–16 NOAA/hazardous weather testbed spring forecasting experiments. *Wea. Forecasting*, **34**, 61–79, https://doi.org/10.1175/WAF-D-18-0024.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blumberg, W. G., K. T. Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPpy: An open-source sounding analysis toolkit for the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **98**, 1625–1636, https://doi.org/10.1175/BAMS-D-15-00309.1.

Brooks, H. E., and C. A. Doswell III, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting*, **11**, 288–303, https://doi.org/10.1175/1520-0434(1996)011<0288:ACOMOA>2.0.CO;2.

Burlingame, B. M., C. Evans, and P. J. Roebber, 2017: The influence of PBL parameterization on the practical predictability of convection initiation during the Mesoscale Predictability Experiment (MPEX). *Wea. Forecasting*, **32**, 1161–1183, https://doi.org/10.1175/WAF-D-16-0174.1.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

——, M. C. Coniglio, B. E. Coffer, G. Thompson, M. Xue, and F. Kong, 2015: Sensitivity of 24-h forecast dryline position and structure to boundary layer parameterizations in convection-allowing WRF Model simulations. *Wea. Forecasting*, **30**, 613–638, https://doi.org/10.1175/WAF-D-14-00078.1.

——, and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1433–1448, https://doi.org/10.1175/BAMS-D-16-0309.1.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Wea. Forecasting*, **30**, 591–612, https://doi.org/10.1175/WAF-D-14-00105.1.

——, ——, ——, ——, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884, https://doi.org/10.1175/WAF-D-16-0193.1.

Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF Model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, https://doi.org/10.1175/WAF-D-12-00103.1.

Davis, C. A., B. G. Brown, and R. G. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

——, ——, and ——, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, https://doi.org/10.1175/MWR3146.1.

Deardorff, J. W., 1972: Numerical investigation of neutral and unstable planetary boundary layers. *J. Atmos. Sci.*, **29**, 91–115, https://doi.org/10.1175/1520-0469(1972)029<0091:NIONAU>2.0.CO;2.

Dowell, D. C., and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.

Flora, M. L., P. S. Skinner, C. K. Potvin, A. E. Reinhart, T. A. Jones, N. Yussouf, and K. H. Knopfmeier, 2019: Object-based verification of short-term, storm-scale probabilistic mesocyclone guidance from an experimental Warn-on-Forecast system. *Wea. Forecasting*, **34**, 1721–1739, https://doi.org/10.1175/WAF-D-19-0094.1.

Gallo, B.T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1.

García-Díez, M., J. Fernández, L. Fita, and C. Yagüe, 2013: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quart. J. Roy. Meteor. Soc.*, **139**, 501–514, https://doi.org/10.1002/qj.1976.

Hong, S., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Hu, X., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844, https://doi.org/10.1175/2010JAMC2432.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Jahn, D. E., and W. A. Gallus, 2018: Impacts of modifications to a local planetary boundary layer scheme on forecasts of the Great Plains low-level jet environment. *Wea. Forecasting*, **33**, 1109–1120, https://doi.org/10.1175/WAF-D-18-0036.1.

Janjić, Z. I., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp., http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf.

Johnson, A., X. Wang, M. Xue, and F. Kong, 2011: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 spring experiment. Part II: Ensemble clustering over the whole experiment period. *Mon. Wea. Rev.*, **139**, 3694–3710, https://doi.org/10.1175/MWR-D-11-00016.1.

——, ——, F. Kong, and M. Xue, 2013: Object-based evaluation of the impact of horizontal grid spacing on convection-allowing forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425, https://doi.org/10.1175/MWR-D-13-00027.1.

——, ——, Y. Wang, A. Reinhart, A. J. Clark, and I. L. Jirak, 2020: Neighborhood- and object-based probabilistic verification of the OU MAP ensemble forecasts during 2017 and 2018 Hazardous Weather Testbeds. *Wea. Forecasting*, **35**, 169–191, https://doi.org/10.1175/WAF-D-19-0060.1.

Jones, T. A., P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith, 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast system using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, https://doi.org/10.1175/BAMS-84-12-1797.

Kerr, C. A., D. J. Stensrud, and X. Wang, 2017: Verification of convection-allowing model ensemble analyses of near-storm environments using MPEX upsonde observations. *Mon. Wea. Rev.*, **145**, 857–875, https://doi.org/10.1175/MWR-D-16-0287.1.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

May, R. M., S. C. Arms, J. R. Leeman, and J. Chastang, 2017: Siphon: A collection of python utilities for accessing remote atmospheric and oceanic datasets. Unidata, accessed 29 July 2019, https://doi.org/10.5065/D6CN72NW.

Mellor, G. L., and T. Yamada, 1974: A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.*, **31**, 1791–1806, https://doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2.

——, and ——, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, https://doi.org/10.1029/RG020i004p00851.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Moeng, C.-H., and P. P. Sullivan, 1994: A comparison of shear- and buoyancy-driven planetary boundary layer flows. *J. Atmos. Sci.*, **51**, 999–1022, https://doi.org/10.1175/1520-0469(1994)051<0999:ACOSAB>2.0.CO;2.

Morris, M. T., J. R. Carley, E. Colón, A. Gibbs, M. S. De Pondeca, and S. Levine, 2020: A quality assessment of the Real-Time Mesoscale Analysis (RTMA) for aviation. *Wea. Forecasting*, **35**, 977–996, https://doi.org/10.1175/WAF-D-19-0201.1.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338, https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.

Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, https://doi.org/10.1023/B:BOUN.0000020164.04146.98.

——, and ——, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020

Olson, J. B., J. S. Kenyon, W. M. Angevine, J. M. Brown, M. Pagowski, and K. Suselj, 2019: A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW. NOAA Tech. Memo. OAR GSD-61, 37 pp., https://doi.org/10.25923/n9wm-be49.

Parker, M. D., 2014: Composite VORTEX2 supercell environments from near-storm soundings. *Mon. Wea. Rev.*, **142**, 508–529, https://doi.org/10.1175/MWR-D-13-00167.1.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. *Mon. Wea. Rev.*, **143**, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

——, K. L. Elmore, and S. J. Weiss, 2010: Assessing the impacts of proximity sounding criteria on the climatology of significant tornado environments. *Wea. Forecasting*, **25**, 921–930, https://doi.org/10.1175/2010WAF2222368.1.

——, and Coauthors, 2019: Systematic comparison of convection-allowing models during the 2017 NOAA HWT Spring Forecasting Experiment. *Wea. Forecasting*, **34**, 1395–1416, https://doi.org/10.1175/WAF-D-19-0056.1.

Powers, J., and Coauthors, 2017: The Weather Research and Forecasting (WRF) Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast system. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stratman, D. R., and K. A. Brewster, 2017: Sensitivities of 1-km forecasts of 24 May 2011 tornadic supercells to microphysics parameterizations. *Mon. Wea. Rev.*, **145**, 2697–2721, https://doi.org/10.1175/MWR-D-16-0282.1.

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, https://doi.org/10.1175/WAF-D-11-00116.1.

Van der Walt, S., J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, 2014: Scikit-image: Image processing in python. *PeerJ*, **2**, e453, https://doi.org/10.7717/peerj.453.

Wade, A. R., M. C. Coniglio, and C. L. Ziegler, 2018: Comparison of near- and far-field supercell inflow environments using radiosonde observations. *Mon. Wea. Rev.*, **146**, 2403–2415, https://doi.org/10.1175/MWR-D-17-0276.1.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/6/2567/4946762/mwrd190389.pdf by NOAA Central Library user on 11 August 2020